

LL(k) 커버링 문법의 확장

(An Extension of LL(k) Covering Grammars)

이 경 옥 [†] 최 광 무 ^{**}

(Gyung-Ok Lee) (Kwang-Moo Choe)

요 약 본 논문에선 LR 문법의 부분 클래스를 동치인 LL 문법으로 변환하는 방법을 제시한다. 이 변환이 적용 가능한 문법을 확장된 k -transformable 문법이라 정의한다. 변환된 문법은 left-to-right 커버의 성질을 만족한다 기존 연구에서 제시한 변환 방법은 LR 문법의 부분 클래스인 k -transformable 클래스와 PLR 클래스를 LL 문법으로 바꾼다. 이 논문에서 제시하는 새로운 변환 방법의 적용 가능한 문법의 범위는 k -transformable 클래스와 PLR 클래스를 포함한다. 기존의 커버링 성질을 만족하는 LL로의 문법 변환은 보편적인 LR 파서의 행동을 시뮬레이션하여 얻어진다. 이 과정에서 쉬프트, 리듀스 행동 이외에 무한의 가능성을 가진 스택 스트링의 유한 표현을 위해 리덕션 심볼에 대한 예상 행동이 추가된다. 본 논문에서는 파싱 문맥을 나타내는 LR 아이템들을 기존의 스택 스트링 표현 형태에 추가하여 스택 스트링 표현법을 정제하고, 리덕션 심볼에 대한 예상 방법을 확장하는 정형식을 제시한다. 이에 근거하여 LL 커버링 문법이 존재하는 클래스를 확장된 k -transformable 문법으로 확장시킨다.

Abstract A new transformation of a subclass of LR(k) grammars into equivalent LL(k) grammars is studied. The subclass of LR(k) grammars is called extended k -transformable. The transformed LL(k) grammars left-to-right cover the original LR(k) grammars. Previous transformations transform k -transformable and PLR(k) subclasses of LR(k) into LL(k). The new transformation is more powerful in that it handles the extended k -transformable subclass of LR(k), which strictly includes k -transformable and PLR(k) classes. The previous covering transformations into LL grammars are obtained by simulating the actions of the conventional LR parser. Specially, a predict action of reduction goals is added to the action set in order to finitely represent stack string. In this paper, the stack string representation is refined by adding LR items to represent a parsing context, and the prediction of reduction goal is extended by generalizing the prediction formalism. Based on them, the previous grammar classes with LL(k) covering grammars are extended to extended k -transformable grammars.

1. 서 론

LR 문법은 왼쪽에서 오른쪽으로 입력 스트링을 읽으면서, 길이 k 의 입력 심볼들을 미리 보는 방식으로 결정적(deterministic)으로 상향식(bottom-up) 파싱을 할 수 있는 가장 큰 범위의 문법이다. LL 문법은 이에 대응되는 하향식(top-down) 파싱 방법이다. 본 논문에선

LR 문법 중에서 커버링 성질을 만족하는 LL 문법으로의 변환이 존재하는 문법에 관심이 있다. LL 커버링 변환이 존재하는 문법 클래스로는 k -transformable 문법과 PLR 문법이 정의되었다. PLR 문법은 k -transformable 문법의 부분 집합이라고 [1]에서 언급되었고, k -transformable 문법은 LL 커버링이 존재하는 문법의 최대 한계 문법일거라고 추측되었다 [2]. 그러나 PLR 문법이면서 k -transformable이 아닌 문법이 발견되었으며 [3], 이는 두 문법 클래스를 포함하는 보다 일반화된 문법 클래스의 정의와 커버링 변환에 대한 연구의 필요성의 동기를 부여하였다.

LL 커버링 변환 문법은 LR 파서의 각 행동을 시뮬레

[†] 비 회 원 : 한국과학기술원 전산학과
golcc@pllab.kaist.ac.kr

^{**} 종신회원 : 한국과학기술원 전산학과 교수
choe@pllab.kaist.ac.kr

논문접수 : 1998년 12월 24일
심사완료 : 1999년 6월 7일

이전하여 얻는다. LR 파서는 스택 심볼을 푸시(push)하거나 헨들에 해당하는 스트링을 팝(pop)하는 방식으로 스택 스트링을 변화시킨다. 이로부터 LL 커버링 문법은 파서의 행동을 시뮬레이션하기 위해서, 스택 스트링에 대한 정보를 포함하고 있어야 한다는 생각을 도출할 수 있다. 한편 스택 스트링의 길이는 무한할 수 없기에, 스택 스트링을 그대로 문법의 구성 요소로 사용할 순 없다. 유한한(finite) 표현법으로 스택 스트링을 나타내기 위해서, 리덕션(reduction) 심볼을 예상하여 이를 문법 심볼로 인코드(encode)하였다. 보편적인 LR 파서에선 헨들이 완전히 스택에 놓여진 후에 리덕션 심볼로 리덕션하게 된다. 하나 경우에 따라선 스택 상위의 특정 스트링이 특정 심볼로 리덕션이 될지를 리덕션 시간 전에 미리 알 수 있다. 스택 스트링 θ 에 대해서, θ 의 상위 부분에 해당하며 길이가 제한된 a 가 A 로 리덕션됨이 예상되면, θ 는 $[A, a]$ 형태로 인코드되어서 표현된다. 이는 LL 커버링 문법의 너터미널을 형성하게 된다.

이 논문에선 PLR(k) 문법과 k -transformable 문법에서 제시한 LL 커버링 방법의 확장을 제안하고자 한다. 기본 생각은 예상이 가능한 스택 스트링과 리덕션 심볼을 찾는 기존 방법에 대한 확장에 근거한다. 이는 두가지로 요약될 수 있다. (i) 과거에 쓰인 스택 스트링의 표현인 $[A, a]$ 형태를 파싱 문맥을 나타내는 q^A, q_a^A 를 덧붙여, $[q^A, a, q_a^A]$ 형태로 정제(refine)한다. (ii) LR 상태의 생성 과정에 관련된 아이템들의 전이를 밀방향(closure) 전이, 옆방향(goto) 전이로 나눌 때 k -transformable 문법에서는 예상이 가능한 리덕션 심볼을 찾기 위해 밀방향 전이의 경로만을 조사하였다. 이에 반해 이 논문에선 밀방향 전이 뿐만 아니라 옆방향 전이의 경로도 같이 조사한다. 이런 두가지 근본 생각은 비결정적인(nondeterministic) LR(k) 유한 상태 오토마톤을 사용하여 정형화된다. 즉 오토마톤의 경로들은 (i)에서 언급한대로 정제된 경로 집합으로 표현되고, 경로들은 (ii)에서의 방법대로 옆방향 전이까지 포함하여 예상할 수 있는 심볼들을 찾게된다. 제안된 (i), (ii)에 근거하여 [1, 2]에서의 방법과 유사하게 LR 파서의 행동은 문법의 규칙으로 표현된다. 이 결과 생성된 문법은 LL 커버링 변환이 되며 이 변환을 적용할 수 있는 문법은 확장된 k -transformable 문법으로 정의된다. 본 논문의 구성은 다음과 같다. 2장에선 본 논문의 진행을 위해 필요한 표기법과 기본 정의들을 정의한다. 3장에선 본 논문에서 제안하는 확장된 예상 방법을 정형적으로 다루고 이를 이용한 LL 문법으로의 변환 방법을 제안하

게 된다. 4장에선 주어진 변환에 대한 LL 커버링 성질을 증명한다. 그리고 5장에서 논문의 결론을 맺는다.

2. 표기법과 기본 정의들

표기법과 기본 정의들은 [4, 5]를 따르며 독자들이 이를 잘 알고 있음을 가정한다.

문법은 G 로 나타내며, 심볼 k 는 입의 고정된 음수가 아닌 정수를 나타낸다. G' 는 [4, pp.372]에서의 G 의 추가된(augmented) 문법이다. G 는 S' 에 의해 추가된 문법을 가정하고 또한 축약된(reduced) 문법 [5]임을 가정한다.

$G_1 = (N_1, \Sigma, P_1, S_1)$ 와 $G_2 = (N_2, \Sigma, P_2, S_2)$ 를 $L(G_1) = L(G_2)$ 이고, h 를 P_2 에서 P_1 으로의 homomorphism이라고 하자. G_2 가 h 상에서 G_1 을 left-to-right 커버한다는 말은 다음의 조건을 만족할 때이다. 만일 $S_2 \Rightarrow_m^r u$ 인 π_2 가 G_2 에 있다면, $S_1 \Rightarrow_m^{h(\pi_2)} u$ 인 $h(\pi_2)$ 가 G_1 에 있고, 만일 $S_1 \Rightarrow_m^r u$ 인 π_1 가 G_1 에 있다면, $h(\pi_1) = \pi_2$ 이며 $S_2 \Rightarrow_m^r u$ 인 π_2 가 G_2 에 있다. 음수가 아닌 k 에 대해서 G_2 가 G_1 을 left-to-right 커버하고, G_2 가 LL(k)이면, G_2 는 G_1 의 LL(k) 커버라고 한다. 본 논문은 LR(k) 문법에 대한 LL(k) 커버에 관심을 두고 있기에, 앞으로 언급하는 G 는 LR(k) 문법임을 가정한다.

k 가 음수가 아닌 정수이고 a 는 스트링이라 하자. k : a 는 $|a| \leq k$ 이면 a 와 동일하고 그외의 경우엔 a 의 처음 k 개의 심볼들과 동일하다. 이때 $FIRST_k(a) = \{x \in G, x \in \Sigma^*\}$ 로 정의된다. $[A \rightarrow a, \beta, u]$ 가 LR(k) 아이템이면 $FIRST_k([A \rightarrow a, \beta, u]) = \{x \in FIRST_k(\beta u)\}$ 로 정의된다. q 가 LR(k) 아이템들의 집합이면, $FIRST_k(q) = \bigcup_{[A \rightarrow a, \beta, u] \in q} FIRST_k([A \rightarrow a, \beta, u])$ 이다.

I_k^G 를 G 상의 LR(k) 아이템들의 집합이라 하자. 비결정적인 LR(k) 유한 상태 오토마톤은 $\widehat{M}_k^G = (I_k^G, V, \rightarrow [S' \rightarrow .S, \$, I_k^G])$ 로 구성된다. 여기서 상태들의 집합은 I_k^G , 전이 심볼들의 집합은 V , 초기 상태는 $[S' \rightarrow .S, \$]$, 최종 상태들의 집합은 I_k^G 이다. \rightarrow 는 $I_k^G \times (V \cup \{\epsilon\})$ 에서 $2^{I_k^G}$ 로의 전이 함수이며 옆방향 전이와 밀방향 전이로 구분된다. (i) 옆방향 전이: $X \in V$ 일 때 $[A \rightarrow a.X\beta, u] \rightarrow^X [A \rightarrow aX.\beta, u]$ 이고 (ii) 밀방향 전이: $v \in FIRST_k(\beta u)$ 일 때 $[A \rightarrow a.B\beta, u] \rightarrow [B \rightarrow .\gamma, v]$ 이다. $a \in V^*$ 에 대해 \rightarrow^* 는 \rightarrow 의 재귀적 전이 닫힘(reflexive transitive

closure) 관계를 나타낸다. 즉 $I_0 \rightarrow^{X_1} I_1 \rightarrow X_2 \dots \rightarrow^{X_n} I_n (n \geq 0)$ 이면 $I_0 \rightarrow^{X_1 \dots X_n} I_n$ 으로 표기한다.

어떤 LR(k) 아이템이 슈프트 또는 리듀스 아이터이면 **터미널(terminal) 아이터**이라 한다. 함수 T는 주어진 아이터으로부터 유도되는 터미널 아이터들을 취하며 $T([A \rightarrow a, \beta, u]) = \{ [B \rightarrow \gamma, \delta, v] \mid [A \rightarrow a, \beta, u] \rightarrow^* [B \rightarrow \gamma, \delta, v], [B \rightarrow \gamma, \delta, v] \text{는 터미널 아이터이다.} \}$ 로 정의된다. q가 LR(k) 아이터들의 집합일 때 $T(q) = \{ [B \rightarrow \gamma, \delta, v] \mid [B \rightarrow \gamma, \delta, v] \in T([A \rightarrow a, \beta, u]), [A \rightarrow a, \beta, u] \in q \}$ 이다. 우리는 여기서 터미널 아이터로부터의 전이는 없음을 관찰할 수 있다.

\widehat{M}_k^C 내의 전이들의 나열 $\Gamma = I_0 \rightarrow^{X_1} I_1 \rightarrow^{X_2} \dots \rightarrow^{X_n} I_n$ 은 **경로(path)**라고 한다. I_n 이 터미널이면 Γ 는 **터미널 경로**라고 한다. 경로내의 밀방향 전이들의 나열 Γ 에 대해 바로 전의 전이와 바로 후의 전이가 존재할 경우에 이들이 밀방향 전이가 아니라는 성질을 만족하면 Γ 은 **세그먼트(segment)**라 한다.

δ 는 $I_k^C \times (V \cup \{\epsilon\})$ 에서 $2^{I_k^C}$ 로의 함수이며 \rightarrow 의 재귀적 전이 닫힘 관계로부터 정의된다. 즉 $[A \rightarrow a, \beta, u] \in I_k^C, X \in V \cup \{\epsilon\}$ 에 대해 $\delta([A \rightarrow a, \beta, u], X) = \{ [B \rightarrow \gamma, \delta, v] \mid [A \rightarrow a, \beta, u] \rightarrow^* [B \rightarrow \gamma, \delta, v] \}$ 이다. q가 LR(k) 아이터들의 집합이면, $\delta(q, X) = \bigcup_{I \in q} \delta(I, X)$ 이고, $X_i \in V, i = 1, \dots, n$ 이면 $\delta(q, X_1 \dots X_n) = \delta(\dots \delta(\delta(q, X_1), X_2), \dots), X_n)$ 이다.

\widehat{M}_k^C 를 결정적인(deterministic) 오토마톤으로 변환한 $M_k^C = (Q, V, \delta^{LR}, q_0^S, \phi)$ 는 LR(k) 오토마톤으로 정의한다. LR(k) 상태들의 집합 Q는 $\{q_0^S\} \cup \{\delta^{LR}(q, X) \mid q \in Q, X \in V\}$, 전이 심볼들은 V, 전이 함수 δ^{LR} 는 $\delta^{LR}(q, X) = \delta(q, X)$, 시작 심볼 q_0^S 는 $\delta(S^* \rightarrow .S, \$)$, 최종 상태들의 집합은 ϕ 으로 정의한다.

$[\theta]_k$ 는 $VALID_k(\theta)$ 를 나타낸다. LR(k) 파서는 M_k^C 로부터 정의된다. LR(k) 파서의 컨피규레이션(configuration)은 $[\epsilon]_k \dots [\theta]_k$ 가 스택 스트링이고 x가 남은 입력 스트링일 때 $[\epsilon]_k \dots [\theta]_k \vdash x$ 형태로 주어진다. 마지막 심볼 $[\theta]_k$ 는 현 상태를 나타낸다. 입력 스트링 x에 대해 초기 컨피규레이션은 $[\epsilon]_k \vdash x$ 이고 최종 컨피규레이션은 $[\epsilon]_k [S]_k \vdash \phi$ 이다.

$[\epsilon]_k \vdash x \Rightarrow^* [\epsilon]_k \dots [\theta]_k \vdash \phi \Rightarrow^* [\epsilon]_k [S]_k \vdash \phi$ 인 컨피규레이션 $[\epsilon]_k \dots [\theta]_k \vdash \phi$ 은 **타당한 LR(k) 파싱 컨피규레이션**이다. 초기 컨피규레이션에서 최종 컨피규레이션으로의 동작들의 배열은 **타당한 LR(k) 파싱 동작들의 배열**이

다. 이 배열의 부분 배열은 **타당한 LR(k) 파싱 동작들의 부분 배열**이다.

파서의 행동은 슈프트 행동과 리듀스 행동으로 나누어지며 각 행동은 슈프트 아이터와 리듀스 아이터에 의해 결정된다. 출력 함수 τ 는 파서의 이동의 집합으로부터 $P \cup \{\epsilon\}$ 로의 함수로 $\tau([\theta]_k \vdash x \Rightarrow [\theta]_k [\theta a]_k \vdash \phi) = \epsilon$ 이고 $\tau([\eta]_k \dots [\eta a]_k \vdash \phi \Rightarrow [\eta]_k [\eta A]_k \vdash \phi) = A \rightarrow a$ 로 정의된다.

q를 LR(k) 아이터들로 구성된 집합이라 할때 q에 속하는 $v \in FIRST_k(\beta u)$ 를 만족하는 터미널 아이터들 $[A \rightarrow a, \beta, u], [B \rightarrow \gamma, \delta, v]$ 에 대해 $[A \rightarrow a, \beta, u]$ 가 슈프트 아이터이면, 이 아이터들은 슈프트/리듀스 충돌을 유발한다고 하고, $[A \rightarrow a, \beta, u]$ 가 리듀스 아이터이면, 이 아이터들은 리듀스/리듀스 충돌을 유발한다고 한다.

앞으로의 남은 장들에서 언급될 LR(k) 파싱 관련 성질 몇 가지를 살펴보고 이 장을 마치고자 한다.

성질 2.1 \widehat{M}_k^C 상에 $[A \rightarrow a, \beta, u] \rightarrow^* [B \rightarrow \gamma, \delta, v]$ 의 전이가 존재하고 어떤 존속할 수 있는 스트링(viable prefix) θ 에 대해 $[A \rightarrow a, \beta, u] \in VALID_k(\theta)$ 이면 필요충분조건으로 G 상에 $a = \beta \gamma, u = kw, v = kzu$ 의 조건 하에 유도 과정 $S \Rightarrow^* \theta A u \Rightarrow^* \theta a w \Rightarrow^* \theta \beta B z w \Rightarrow^* \theta \beta \gamma \delta z w \Rightarrow^* \theta \beta \gamma z u$ 가 존재한다. □

성질 2.2 G 상에 $S \Rightarrow^* \theta A u \Rightarrow^{r_1} \theta a w \Rightarrow^{r_2} \theta \beta B z w \Rightarrow^{r_3} \theta \beta \gamma \delta z w \Rightarrow^{r_4} \theta \beta \gamma z u$ 인 문법 규칙 스트링인 $\widetilde{r}_1, \widetilde{\pi}_1, \widetilde{r}_2, \widetilde{\pi}_2$ 가 존재하기 위한 필요충분조건으로 $\tau(\pi_2) = \widetilde{\pi}_2^R, \tau(r_2) = \widetilde{r}_2, \tau(\pi_2) = \widetilde{\pi}_2^R, \tau(r_1) = \widetilde{r}_1$ 인 조건 하에 타당한 LR(k) 파싱 동작들의 부분 배열인 $[\theta]_k \dots [\theta \beta \gamma]_k \vdash yzw \Rightarrow^{r_1} [\theta]_k \dots [\theta \beta \gamma \delta]_k \vdash zu \Rightarrow^{r_2} [\theta]_k \dots [\theta \beta B]_k \vdash zu \Rightarrow^{r_3} [\theta]_k \dots [\theta a]_k \vdash w \Rightarrow^{r_4} [\theta]_k [\theta A]_k \vdash w$ 이 있다. □

성질 2.1과 성질 2.2로부터 다음의 성질을 얻을 수 있다.

성질 2.3 $[A \rightarrow a, \beta, u] \rightarrow^* [B \rightarrow \gamma, \delta, v]$ 의 전이가 \widehat{M}_k^C 상에 있고 어떤 존속할 수 있는 스트링 θ 에 대해 $[A \rightarrow a, \beta, u] \in VALID_k(\theta)$ 이면 필요충분조건으로 $a = \beta \gamma, u = kw, v = kzu$ 인 조건 하에 타당한 LR(k) 파싱 동작들의 부분 배열인 $[\theta]_k \dots [\theta \beta \gamma]_k \vdash yzw \Rightarrow [\theta]_k \dots [\theta \beta \gamma \delta]_k \vdash zu \Rightarrow [\theta]_k \dots [\theta \beta B]_k \vdash zu \Rightarrow [\theta]_k \dots [\theta a]_k \vdash w \Rightarrow [\theta]_k [\theta A]_k \vdash w$ 이 있다. □

3. LL 문법의 생성

이 장에선 \widehat{M}_k^C 의 경로 집합의 표현법을 정의한다.

정의된 경로 집합은 LR 파서의 각 컨피규레이션을 묘사하는데 사용되며 본 논문에서 제안하는 변환 문법의 언터미널 심볼에 대응된다. 모든 가능한 컨피규레이션을 묘사할 수 있는 유한한 경로 집합들의 컬렉션(collection)을 얻기 위한 예상 가능한 리덕션 심볼과 스택 스트링을 찾는 관계를 정의하게 된다. LR 파서의 행동들에 따른 컨피규레이션의 변화를 문법의 규칙으로 표현하여 LL 문법을 얻게 된다.

3.1 경로들의 집합

$A \in N$ 일 때 q^A 는 $[A \rightarrow \omega, u]$ 형태의 아이템들로 구성된 LR(k) 상태의 부분 집합이다. $a \in V^*$ 이고 $\widehat{q^A} = \delta(q^A, a)$ 일 때 q_a^A 는 $q_a^A \subseteq T(\widehat{q^A})$ 를 만족하는 아이템들의 집합이라 하자. 이 때 $\langle q^A, a, q_a^A \rangle$ 는 $a = X_1 X_2 \dots X_n, X_i \in V, i=1$ 라 할 때 $\{Y|Y = Y_0 \rightarrow^{X_1} Y_1 \rightarrow^{X_2} \dots \rightarrow^{X_n} Y_n$ 이고 $i=0, 1, \dots, n$ 대해 $Y_i = I_{i,0} \rightarrow I_{i,1} \rightarrow \dots \rightarrow I_{i,m_i}$ 는 터미널 경로이다. 또한 $I_{0,0} \in q^A$ 이고 $I_{n,m_n} \in q_a^A$ 이다.)로 정의된다. 여기서 각 Y_i 는 세그먼트이다. Y_i 를 Y 의 $segment_i$ 라 한다. $a = \zeta\delta$ 일 때 $SEG_{|\delta|} \langle q^A, a, q_a^A \rangle = \{I_0, I_1, \dots, I_m | I_0 \rightarrow I_1 \rightarrow \dots \rightarrow I_m$ 는 $\langle q^A, a, q_a^A \rangle$ 내의 임의의 경로의 $segment_{|\delta|}$ 이다.)이다. 특히 $SEG_{|\delta|} \langle q^A, a, q_a^A \rangle$ 는 $LSEG \langle q^A, a, q_a^A \rangle$ 로 표기하기도 한다. $\langle q^A, a, q_a^A \rangle$ 에 속하는 경로를 그림 1에서 보여준다.

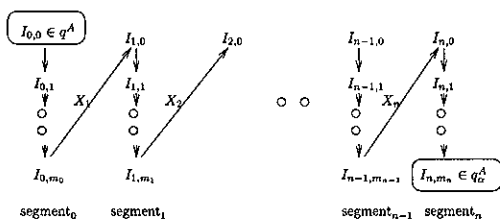


그림 1 $a = X_1 \dots X_n$ 일 때 $\langle q^A, a, q_a^A \rangle$ 내의 경로

타당한 LR(k) 파싱 컨피규레이션 $[\epsilon]_k \dots [\theta]_k \dots [\theta\alpha]_{|\delta|}$ 가 아래 조건들을 모두 만족하면 $\langle q^A, a, q_a^A \rangle$ 에 의해 묘사된다고 한다. 즉 어떤 $[B \rightarrow \gamma, \delta, v] \in q_a^A$ 에 대해 다음행동은 $[B \rightarrow \gamma, \delta, v]$ 에 의해 결정되며, 타당한 LR(k) 파싱 동작들의 부분 배열인 $[\epsilon]_k \dots [\theta]_k \dots [\theta\alpha]_{|\delta|} \rightsquigarrow^* [\epsilon]_k \dots [\theta]_k \dots [\theta\omega]_{|\delta|} \rightsquigarrow [\epsilon]_k \dots [\theta]_k \dots [\theta A]_{|\delta|} z(x = \gamma z, u = k z)$ 이 존재한다. 임의의 타당한 LR(k) 파싱 컨피

규레이션 $[\epsilon]_k \dots [\theta]_k \dots x$ 는 q_a^S 를 $\delta(q^S, \theta)$ 인 아이тем들의 집합이라 할 때 $\langle q^S, \theta, T(q_a^S) \rangle$ 에 의해 묘사된다. 한편 경로 집합의 표현법은 스택 스트링을 포함하기에 무한하다. 이 경로 집합을 문법 심볼로 사용하기 위해 모든 타당한 LR(k) 파싱 컨피규레이션을 묘사할 수 있는 유한한 경로 집합들의 집합(collection)이 존재해야 한다. 이 문제는 이미 언급한 바와 같이 스택 스트링과 리덕션 심볼을 미리 찾음으로 해결하고자 한다.

3.2 리덕션 심볼에 대한 예상

리덕션 심볼을 예상하기 위해서 주어진 경로 집합내의 각 경로에 대해 특정 세그먼트 상에 특정 심볼의 존재를 확인하게 된다. 다음 관계는 이를 정의한다.

정의 3.1 (\mathbb{I} 관계) q^A, a, q_a^A 는 앞서의 정의와 동일하고 $B \in N, a = \beta\gamma$ 라 하자. $\langle q^A, a, q_a^A \rangle, B, \gamma \in \mathbb{I}$ 가 되기 위해선 $\langle q^A, a, q_a^A \rangle$ 내의 각 경로의 $segment_{|\delta|}$ 인 $Y, Y = I_0 \rightarrow \dots \rightarrow I_i \rightarrow \dots \rightarrow I_n$ 에 대해 $I_i = [B \rightarrow \phi, v]$ 인 i 가 존재하여야 한다. □

위의 관계를 만족하는 B 를 찾기 위해서, 각 세그먼트들은 많아야 한번 방문하는 것으로 충분하다. \mathbb{I} 관계를 경로 집합을 나타내는 표현법에 맞추어 재현하여 아래와 같은 관계를 정의한다.

정의 3.2 (\mathbb{II} 관계) $\langle q^A, a, q_a^A \rangle, B, \gamma \in \mathbb{II}, a = \beta\gamma$ 이고 q_a^A, q_a^B 는 $q_a^A \supseteq q_a^B, q_a^B = \overline{q_a^A}$ 를 만족하는 아이тем들의 집합이라 하자. 이때 $\langle q^A, a, q_a^A \rangle \mathbb{II} \langle q^B, \gamma, q_a^B \rangle, \langle q^A, \beta B, q_a^{\beta B} \rangle$ 가 성립하기 위해선 아래의 조건을 만족해야 한다. 즉 $\langle q^A, a, q_a^A \rangle$ 내의 각 경로의 $segment_{|\delta|}, Y = I_0 \rightarrow I_1 \rightarrow \dots \rightarrow I_n$ 에 대해 $I_i = [B \rightarrow \phi, v], q_a^B = \bigcup I_i, \overline{q_a^{\beta B}} = \bigcup \{[C \rightarrow \xi B, \delta, w] | I_{i-1} = [C \rightarrow \xi, B\delta, w]\}$ 로 정의할 때 $q_a^{\beta B} = T(\overline{q_a^{\beta B}})$ 인 $i, 1 \leq i \leq n$ 가 존재한다. (여기서 Y 는 $\langle q^A, a, q_a^A \rangle$ 내의 모든 경로들 상에서 정의된다.) □

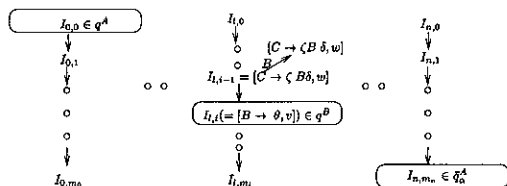


그림 2 | β | = 1

그림 2는 정의 3.2에서의 아이тем들의 상황을 보여주

q_a^A 는 루프 p_0, p_1, \dots, p_m 의 세 번 이상의 발생을 포함하지 않는다. (배열 $SEG_0 \langle q^A, \bar{a}, q_a^A \rangle, SEG_1 \langle q^A, \bar{a}, q_a^A \rangle, \dots, SEG_{|a|} \langle q^A, \bar{a}, q_a^A \rangle$ 는 $SEG_0 \langle q^A, \alpha, q_a^A \rangle, SEG_1 \langle q^A, \alpha, q_a^A \rangle, \dots, SEG_{|a|} \langle q^A, \alpha, q_a^A \rangle$ 에서 루프 p_0, p_1, \dots, p_m 의 $r-1$ 번의 발생을 제거하여 얻어진다; 다음 예제는 이에 대한 한 예를 보인다.) \square

예제 2 G1에 대해 다음의 순환적 경로 집합을 고려해 보자.

$$\left\langle \left\{ [S' \rightarrow .S, \$] \right\}, aaa^3aa, \left\{ \begin{array}{l} [A \rightarrow a.b, d] \\ [A \rightarrow .ab, d] \\ [A \rightarrow .aA, d] \\ [A \rightarrow .bc, d] \end{array} \right\} \right\rangle_{0,1,2,2,2,2,3,4}$$

(여기서 첨자는 그림 4내의 상태번호를 지시한다.)
이 때

$$\left(\left\{ [S' \rightarrow .S, \$] \right\}, aaa^2aa, \left\{ \begin{array}{l} [A \rightarrow a.b, d] \\ [A \rightarrow .ab, d] \\ [A \rightarrow .aA, d] \\ [A \rightarrow .bc, d] \end{array} \right\} \right) \overset{\Pi}{\Pi} \left(\left\{ [A \rightarrow a.b, d] \right\}, \left\{ [A \rightarrow .ab, d] \right\}, \left\{ [A \rightarrow .aA, d] \right\}, \left\{ [A \rightarrow .bc, d] \right\} \right)$$

가 성립한다. 여기서 우리는

$$\left(\left\{ [S' \rightarrow .S, \$] \right\}, aaa^2aa, \left\{ \begin{array}{l} [A \rightarrow a.b, d] \\ [A \rightarrow .ab, d] \\ [A \rightarrow .aA, d] \\ [A \rightarrow .bc, d] \end{array} \right\} \right) \overset{\Pi}{\Pi} \left(\left\{ [A \rightarrow a.b, d] \right\}, \left\{ [A \rightarrow .ab, d] \right\}, \left\{ [A \rightarrow .aA, d] \right\}, \left\{ [A \rightarrow .bc, d] \right\} \right)_{0,1,2,2,3,4} \quad (5)$$

가 성립함을 관찰할 수 있다. 이때 관계(5)의 왼쪽의 경로 집합이 비순환적임을 알 수 있다. \square

성질 3.2의 문맥에서 만일 $\langle \langle q^A, \bar{a}, q_a^A \rangle, B, \bar{\gamma} \rangle \in \mathcal{N}$ 가 성립하지 않는다면 이것은 $\langle \langle q^A, \alpha, q_a^A \rangle, B, \bar{\gamma} \rangle \in \mathcal{N}$ 도 역시 성립하지 않음을 의미한다. 이 과정을 되풀이하여 유용한 성질을 얻을 수 있다. 즉 순환적인 $\langle q^A, \alpha, q_a^A \rangle$ 에 대해 $\langle \langle q^A, \alpha, q_a^A \rangle, B, \bar{\gamma} \rangle \in \mathcal{N}$ 의 성립 여부는 비순환적인 $\langle q^A, \bar{a}, q_a^A \rangle$ 에 대한 $\langle \langle q^A, \bar{a}, q_a^A \rangle, B, \bar{\gamma} \rangle \in \mathcal{N}$ 의 성립 여부에 의해 결정된다. 결과적으로 순환적인 경로 집합에 대해 \mathcal{N} 의 조건을 확인할 필요가 없음을 알 수 있다. 이것은 \mathcal{N} 가 효과적으로 계산될 수 있음을 의미한다 유사하게 \mathcal{N} 관계 역시 효과적으로 계산할 수 있음을 보일 수 있다

3.4 변환

변환의 LL 성질을 위해 \mathcal{N} 관계의 부분 관계인 Π_{LL} 을 정의한 후에 이를 이용한 문법 변환을 제시한다.

정제 3.3 (Π_{LL} 관계) Π_{LL} 은 Condition 1, Condition 2(a), Condition 2(b)를 만족하는 \mathcal{N} 의 부분 관계이다. Condition 1: 두개의 다른 관계 $\langle q^A, \alpha, q_a^A \rangle \Pi_{LL} \langle q^B, \gamma, q_\gamma^B \rangle, \langle q_A, \beta B, q_{\beta B}^A \rangle, \langle q_A, \alpha, q_a^A \rangle \Pi_{LL} \langle q^C, \delta, q_\delta^C \rangle, \langle q_A, \zeta C, q_{\zeta C}^A \rangle$ 에 대해 $FIRST_k(q_\gamma^B) \cap FIRST_k(q_\delta^C) = \emptyset$ 이다.

Condition 2: 모든 $\langle q^A, \alpha, q_a^A \rangle \Pi_{LL} \langle q^B, \gamma, q_\gamma^B \rangle, \langle q^A, \beta B, q_{\beta B}^A \rangle$ 에 대해 다음 조건 (a)와 (b)를 만족한다.

- (a) $FIRST_k(q_\gamma^B) \cap FIRST_k(q_a^A - q_\gamma^B) = \emptyset$.
- (b) q_β^A 가 $T(\delta(q^B, B))$ 일 때에 $FIRST_k(q_{\beta B}^A) \cap FIRST_k(q_\beta^A) = \emptyset$. \square

어떤 문법 G가 어떤 양수 k에 대해 확장된 k-transformable인지의 여부는 다음의 알고리즘으로 결정될 수 있다.

알고리즘 0 (확장된 k-transformable 문법) 어떤 음이 아닌 정수 k에 대해 G를 LR(k)라 하자. $\Pi \subseteq \Pi_{LL}$ 인 Π 에 대해서 알고리즘 1의 결과 순환적인 너터미널을 포함하지 않는 $T(G, \Pi)$ 가 정의되면, $k \geq 1$ 이면 $k' = k$ 이고 $k=0$ 이면 $k'=1$ 인 k' 에 대해 G는 확장된 k' -transformable이다. \square

다음은 G와 Π 가 주어진 경우에 변환 문법 $T(G, \Pi)$ 을 생성하거나, 순환적 너터미널이 $T(G, \Pi)$ 에 존재함을 알리는 알고리즘이다.

알고리즘 1 ($T(G, \Pi)$ 의 생성)

- 입력: G와 Π 관계
- 출력: $T(G, \Pi)$ 또는 오류 메시지
- 방법: $T(G, \Pi)$ 는 $(N, \Sigma \cup \{\epsilon\}, P, S)$ 로 구성되어 있다. 여기서 $q^S = \{[S' \rightarrow .S, \$]\}$ 일 때 $S = \{q^S, \epsilon, T(q^S)\}$, N은 P에 나타나는 너터미널들의 집합, P는 다음 과정에서 얻어지는 문법 규칙들의 집합이다.
 1. $N = \{S\}$
 2. $P = \emptyset$

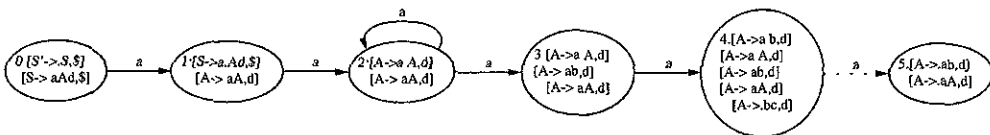


그림 4 관계 (4), (5)상의 SEGs

3. repeat

for $[q^A, a, q^A] \in N$ do

(a) if $a = \text{Athen}$

예상 마침 타입: $P = P \cup \{[q^A, A, q^A] \rightarrow \varepsilon\}$ endif

(b) for $I \in q_a^A$ do

(i) if $\langle q^A, a, q_a^A \rangle \Pi \langle q^B, \gamma, q_\gamma^B \rangle, \langle q^A, \beta B, q_{\beta B}^A \rangle$ 이

고 $I \in q_\gamma^B$ then

예상 타입: $P = P \cup \{[q^A, a, q_a^A] \rightarrow [q^B, \gamma, q_\gamma^B]$
 $[q^A, \beta B, q_{\beta B}^A]\}$

(ii) else if I 가 쉬프트 아이템이고 $[B \rightarrow \gamma, a\delta, v]$ 형태이다 then

읽기 타입: $q_{aa}^A = T(\{[B' \rightarrow \gamma' a, \delta', v'] \mid [B' \rightarrow \gamma' a\delta', v'] \in q_a^A\})$ 일때

$P = P \cup \{[q^A, a, q_a^A] \rightarrow a[q^A, a, q_{aa}^A]\}$

(iii) else if I 가 리듀스 아이템이고 $[B \rightarrow \gamma, v]$ 형태이다 then

리덕션 타입: $a = \beta\gamma$ 이고

$q_{\beta B}^A = T(\{[C \rightarrow \zeta B, \delta, w] \mid ([A \rightarrow \omega, w] \rightarrow^* [C \rightarrow \zeta B\delta, w] \rightarrow^* [B \rightarrow \gamma, v']) \in \langle q^A, a, q_a^A \rangle\})$

일때

$P = P \cup \{[q^A, a, q_a^A] \rightarrow [q^A, \beta B, q_{\beta B}^A]\}$

endif

endfor

(c) 새로이 추가된 문법 규칙에 새로이 나타난 너터미널 심볼은 N 에 더해진다.

endfor

until (N 이 변하지 않거나 순환적 너터미널이 N 에서 발견되었다)

4. if N 에 순환적 너터미널이 없다 then $T(G, M)$ 를 리턴한다

else 순환적 너터미널이 존재함을 알리는 오류 메시지를 리턴한다 end if □

이 후에 나오는 문법은 어떤 양수 k 에 대해 G 는 확장된 k -transformable이고 $T(G, M)$ 는 변환 후 얻어지는 문법임을 가정한다.

각 문법 규칙의 왼쪽은 파싱 시에 현재의 컨피규레이션을 묘사하는 경로 집합에 해당된다. 읽기 타입과 리덕션 타입의 각 규칙은 $LR(k)$ 파서의 행동들을 시뮬레이션하며, 이들 규칙의 오른쪽의 너터미널에 대응되는 경로 집합은 각 행동 수행 후의 컨피규레이션을 묘사하게

된다. 예상 타입의 규칙은 현재의 컨피규레이션을 묘사하는 경로 집합에 대한 무한성을 막기 위해서 유한한 표현법으로 바꾸는 역할을 한다. 예상 마침 타입의 각 규칙은 예상한 대로 리덕션이 일어났음을 의미한다. 기존의 연구와 비교하자면 [1,2]에서의 문법 규칙은 예상 마침 타입: $[A, a] \rightarrow \varepsilon$, 예상 타입: $[A, a] \rightarrow [B, \gamma]$, $[A, \beta B]$, 읽기 타입: $[A, a] \rightarrow a[A, aa]$, 리덕션 타입: $[A, a] \rightarrow [A, \beta B]$ 형태이다.

특히 예상 타입의 규칙은 다음과 같이 제한적으로 사용되고 있다. 즉 [2]에서 제안된 정형론은 $\gamma = \varepsilon$ 인 $[A, a] \rightarrow [B, \varepsilon]$, $[A, aB]$ 형태의 예상 타입만이 가능하다. 그리고 [1]에서 제안된 것은 γ 의 길이가 1 또는 0만이 가능하며, 어떤 δ 에 대해서 $B \rightarrow \gamma\delta$ 의 규칙이 존재해야만 한다. 이에 반해서 본 논문에서 제시된 문법 규칙은 예상 타입의 γ 부분에 대한 제약이 없다. 또한 본 논문에서 제시한 문법의 너터미널은 기존의 $[A, a]$ 형태의 너터미널에 대해 q^A, q_a^A 을 추가하여 $[q^A, a, q_a^A]$ 형태로 만들어 파싱 문맥 상황을 정제한 효과를 보고 있다. 이로서 기존의 제시된 문법 변환이 본 논문의 제시된 문법 변환의 특정한 경우임을 알수 있다. 또한 기존의 문법 변환으로 LL문법을 만들 수 없어도, 제시된 변환은 LL문법을 만들 수 있는 경우를 쉽게 찾을 수 있다.

$[q^A, a, q_a^A]$ 의 q_a^A 는 $LR(k)$ 상태의 부분 집합이다. 따라서 q_a^A 에서 쉬프트와 리듀스 충돌 또는 리듀스와 리듀스 충돌을 유발하는 아이템은 존재하지 않는다.

정리 3.1 G 를 양수 k 에 대해서 $LR(k)$ 이면 $T(G, M)$ 는 $LL(k)$ 이다.

(증명) 정형적이고 자세한 증명 과정은 [3]을 참조하기 바라며, 여기선 기본 아이디어만을 말한다. G 가 $LR(k)$ 문법이며 읽기, 리덕션 타입의 규칙이 $LR(k)$ 행동을 시뮬레이션했기에 이들 타입의 규칙으로부터 LL 문법이 아니게 하는 요인은 발생하지 않는다. 한편 Π_{LL} 관계는 예상 타입과 예상 마침 타입의 규칙에 대한 다른 타입의 규칙과의 LL 문법을 위한 충돌을 야기하는 것을 배제하는 역할을 하기에, 이들 규칙으로 인한 LL 문법이 아니도록 하는 요인 역시 존재할 수 없다. □

한편 G 가 $LR(0)$ 이면 $T(G, M)$ 는 $LL(1)$ 문법이 된다. 이는 정리 3.1과 유사하게 증명될 수 있다.

예제 3 G_1 에 대해 n 관계를 예제 1에서의 \hat{n} 관계 (1)와 (2)로 정의하자. 이때 변환된 문법인 $T(G_1, M)$ 는 순환적인 너터미널을 가지고 있지 않다. 따라서 G_1 은 확장된 1-transformable이다. 한편 G_1 은 [2]에 의해서

$$\begin{aligned}
 r_1^1 &= [\{ [S' \rightarrow S, \$] \}, \epsilon, \{ [S \rightarrow aAd, \$] \}] \rightarrow a [\{ [S' \rightarrow S, \$] \}, a, \left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow aA, d] \\ [A \rightarrow bc, d] \end{array} \right\}] \\
 r_2^1 &= [\{ [S' \rightarrow S, \$] \}, a, \left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow aA, d] \\ [A \rightarrow bc, d] \end{array} \right\}] \rightarrow b [\{ [S' \rightarrow S, \$] \}, ab, \{ [A \rightarrow bc, d] \}] \\
 r_3^1 &= [\{ [S' \rightarrow S, \$] \}, a, \left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow aA, d] \\ [A \rightarrow bc, d] \end{array} \right\}] \rightarrow a [\{ [S' \rightarrow S, \$] \}, aa, \left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow aA, d] \\ [A \rightarrow bc, d] \end{array} \right\}] \\
 r_4^1 &= [\{ [S' \rightarrow S, \$] \}, ab, \{ [A \rightarrow bc, d] \}] \rightarrow c [\{ [S' \rightarrow S, \$] \}, abc, \{ [A \rightarrow bc, d] \}] \\
 r_5^1 &= [\{ [S' \rightarrow S, \$] \}, aa, \left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow aA, d] \\ [A \rightarrow bc, d] \end{array} \right\}] \rightarrow [\{ [A \rightarrow ab, d] \}, \epsilon, \left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow aA, d] \end{array} \right\}] \\
 &\quad [\{ [S' \rightarrow S, \$] \}, aaA, \{ [A \rightarrow aA, d] \}] \\
 r_6^1 &= [\{ [S' \rightarrow S, \$] \}, aa, \left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow aA, d] \\ [A \rightarrow bc, d] \end{array} \right\}] \rightarrow b [\{ [S' \rightarrow S, \$] \}, aab, \left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow bc, d] \end{array} \right\}] \\
 r_7^1 &= [\{ [S' \rightarrow S, \$] \}, abc, \{ [A \rightarrow bc, d] \}] \rightarrow [\{ [S' \rightarrow S, \$] \}, aA, \{ [S \rightarrow aAd, \$] \}] \\
 r_8^1 &= [\{ [S' \rightarrow S, \$] \}, aA, \{ [S \rightarrow aAd, \$] \}] \rightarrow d [\{ [S' \rightarrow S, \$] \}, aAd, \{ [S \rightarrow aAd, \$] \}] \\
 r_9^1 &= [\{ [S' \rightarrow S, \$] \}; aAd, \{ [S \rightarrow aAd, \$] \}] \rightarrow [\{ [S' \rightarrow S, \$] \}, S, \{ [S' \rightarrow S, \$] \}] \\
 r_{10}^1 &= [\{ [S' \rightarrow S, \$] \}, aab, \left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow bc, d] \end{array} \right\}] \rightarrow [[[S' \rightarrow S, \$] \}, aA, \{ [S \rightarrow aAd, \$] \}] \\
 r_{11}^1 &= [\{ [S' \rightarrow S, \$] \}, aab, \left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow bc, d] \end{array} \right\}] \rightarrow c [\{ [S' \rightarrow S, \$] \}, aabc, \{ [A \rightarrow bc, d] \}] \\
 r_{12}^1 &= [\{ [S' \rightarrow S, \$] \}, aabc, \{ [A \rightarrow bc, d] \}] \rightarrow [[[S' \rightarrow S, \$] \}, aaA, \{ [A \rightarrow aA, d] \}] \\
 r_{13}^1 &= [\{ [S' \rightarrow S, \$] \}, aaA, \{ [A \rightarrow aA, d] \}] \rightarrow [[[S' \rightarrow S, \$] \}, aA, \{ [S \rightarrow aAd, \$] \}] \\
 r_{14}^1 &= [\left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow aA, d] \end{array} \right\}, \epsilon, \left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow aA, d] \end{array} \right\}] \rightarrow a \left[\left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow aA, d] \end{array} \right\}, a, \left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow aA, d] \\ [A \rightarrow bc, d] \end{array} \right\} \right] \\
 r_{15}^1 &= [\left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow aA, d] \end{array} \right\}, a, \left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow aA, d] \\ [A \rightarrow bc, d] \end{array} \right\}] \rightarrow [\left[\left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow aA, d] \end{array} \right\}, \epsilon, \left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow aA, d] \end{array} \right\} \right] \\
 &\quad \left[\left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow aA, d] \end{array} \right\}, aA, \{ [A \rightarrow aA, d] \} \right] \\
 r_{16}^1 &= [\left[\left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow aA, d] \end{array} \right\}, a, \left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow aA, d] \\ [A \rightarrow bc, d] \end{array} \right\} \right] \rightarrow b \left[\left[\left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow aA, d] \end{array} \right\}, ab, \left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow bc, d] \end{array} \right\} \right] \right] \\
 r_{17}^1 &= [\left[\left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow aA, d] \end{array} \right\}, aA, \{ [A \rightarrow aA, d] \} \right] \rightarrow [\left[\left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow aA, d] \end{array} \right\}, A, \emptyset \right] \\
 r_{18}^1 &= [\left[\left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow aA, d] \end{array} \right\}, ab, \left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow bc, d] \end{array} \right\} \right] \rightarrow c \left[\left[\left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow aA, d] \end{array} \right\}, abc, \{ [A \rightarrow bc, d] \} \right] \right] \\
 r_{19}^1 &= [\left[\left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow aA, d] \end{array} \right\}, ab, \left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow bc, d] \end{array} \right\} \right] \rightarrow [\left[\left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow aA, d] \end{array} \right\}, A, \emptyset \right] \\
 r_{20}^1 &= [\left[\left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow aA, d] \end{array} \right\}, abc, \{ [A \rightarrow bc, d] \} \right] \rightarrow [\left[\left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow aA, d] \end{array} \right\}, aA, \{ [A \rightarrow aA, d] \} \right] \\
 r_{21}^1 &= [\left[\left\{ \begin{array}{l} [A \rightarrow ab, d] \\ [A \rightarrow aA, d] \end{array} \right\}, A, \emptyset \right] \rightarrow \epsilon \\
 r_{22}^1 &= [\{ [S' \rightarrow S, \$] \}, S, \{ [S' \rightarrow S, \$] \}] \rightarrow [\{ [S' \rightarrow S, \$] \}, S', \emptyset] \\
 r_{23}^1 &= [\{ [S' \rightarrow S, \$] \}, S', \emptyset] \rightarrow \epsilon
 \end{aligned}$$

non-1-transformable이다. 본 논문에서의 연구와 [2]에서의 연구의 비교를 위해선 [3,Appendix A]을 참조하기 바란다. 다음은 $T(GI, M)$ 의 문법 규칙들을 보여주고 있다. $T(GI, M)$ 이 LL(1)임을 판할 수 있다. \square

위의 예제의 M^G 는 그림 5에서 보인다. 주어진 입력 스트링 $aaaba$ 에 대한 GI의 LR(1) 파서는 다음의 움직임 π 를 가진다.

$$[\epsilon]_1 | aaaba \xrightarrow{shift} [\epsilon]_1 [a] | aab\$$$

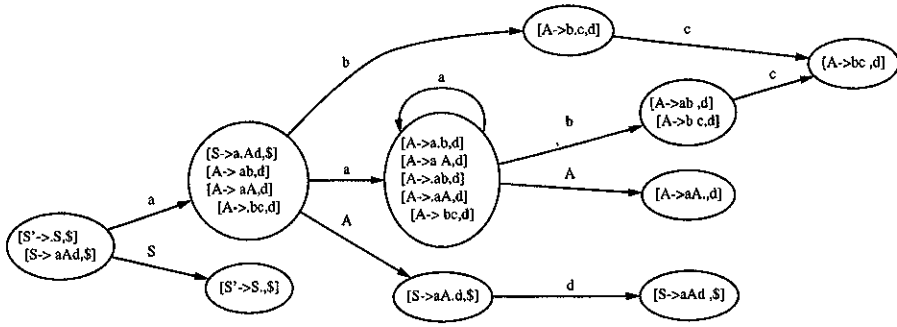


그림 5 M_1^G

$$\begin{aligned} &\Rightarrow \text{shift} [\epsilon]_1 [a]_1 [aa]_1 [abab] \\ &\Rightarrow \text{shift} [\epsilon]_1 [a]_1 [aa]_1 [aaa]_1 [bc] \Rightarrow \text{shift} [\epsilon]_1 [a]_1 \\ & [aa]_1 [aaa]_1 [aaab]_1 \Rightarrow \text{reduce } A \rightarrow ab \\ & [\epsilon]_1 [a]_1 [aa]_1 [aaa]_1 \Rightarrow \text{reduce } A \rightarrow aA [\epsilon]_1 [a]_1 [aA]_1 \\ & \Rightarrow \text{shift} [\epsilon]_1 [a]_1 [aA]_1 [aAd]_1 \\ & \Rightarrow \text{reduce } S \rightarrow aAd [\epsilon]_1 [S]_1 \$ \end{aligned}$$

한편 "aaaba"를 생성하는 $T(G, I)$ 에서의 좌측 유도는 $r_1^1 r_2^3 r_3^5 r_4^7 r_5^9 r_6^{11} r_7^{13} r_8^{15} r_9^{17} r_{10}^{19} r_{11}^{21} r_{12}^{23} r_{13}^{25} r_{14}^{27} r_{15}^{29}$ 라고 하자이다. 여기서 유도 과정 π_T 이 π 의 행동을 시뮬레이션하고 있음을 판측할 수 있다. π 내의 각 컨피규레이션은 π_T 에서 보이는 너트미널 심볼에 대응하는 경로 집합에 대응한다.

가령 컨피규레이션 $[\epsilon]_1 [aaabab]_1$ 은 $\langle [S' \rightarrow S, \$], \epsilon, S \rightarrow aAd, \$ \rangle$ 에 의해 묘사되고 컨피규레이션 $[\epsilon]_1 [a]_1 [aa]_1 [abab]_1$ 은

$$\left\langle \left\{ [S' \rightarrow S, \$] \right\}, aa, \left\{ \begin{array}{l} [A \rightarrow a.b, d] \\ [A \rightarrow .ab, d] \\ [A \rightarrow .aA, d] \\ [A \rightarrow .bc, d] \end{array} \right\} \right\rangle \text{와}$$

$$\left\langle \left\{ \begin{array}{l} [A \rightarrow .ab, d] \\ [A \rightarrow .aA, d] \end{array} \right\}, \epsilon, \left\{ \begin{array}{l} [A \rightarrow .ab, d] \\ [A \rightarrow .aA, d] \end{array} \right\} \right\rangle$$

에 의해 묘사된다.

π_T 에서 $[q^A, \bar{a}, q^{\bar{A}}] \rightarrow [q^{\bar{A}}, \bar{\beta}B, q^{\bar{A}/\bar{\beta}B}]$ 형태 문법 규칙들을 적용 순서대로 나열하면 아래와 같다.

이때 $\tau_T(\pi_T) = \tau_T(r_1^{19} r_2^{13} r_3^9 r_4^{22}) = (A \rightarrow ab)(A \rightarrow aA)(S \rightarrow aAd)(S \rightarrow S)$ 이 성립한다. 한편 "aaaba"를 생성하는 G 의 우측 유도 과정은 $S \Rightarrow_{m1}^r S \Rightarrow_{m2}^r aAd \Rightarrow_{m3}^r aaAa \Rightarrow_{m4}^r aaaba$ 이다.

($r^4 = S' \rightarrow S, r^3 = S \rightarrow aAd, r^2 = A \rightarrow aA, r^1 = A \rightarrow ab$) 여기서 $\tau_T(\pi_T) = \pi^R (\pi = r^4 r^3 r^2 r^1)$ 이다. 사실, $T(G, I)$ 는 G 를 left-to-right 커버한다.

4. LL(k) 커버링 성질

이 장에서는 앞 장에서 제안한 변환 방법이 커버링 성질을 만족함을 보이도록 한다. 앞 장에서 이미 $T(G, I)$ 가 LL 문법임을 보였다. 여기서 G 의 유도 과정과 $T(G, I)$ 의 유도 과정에 대한 대응 관계를 보임으로 증명을 하고자 한다.

증명을 위해서 우선 푸시다운 오토마톤 \widehat{M}_k^G 를 정의한다. 이 오토마톤은 LR(k) 아이템의 관점에서 LR(k) 파서의 행동을 시뮬레이션하기 위해서 정의된다. \widehat{M}_k^G 는 $(I_k^G, \mapsto, [S' \rightarrow S, \$], [S' \rightarrow S, \$])$ 으로 구성된다. 여기서 스택 심볼들의 집합은 I_k^G , 초기 상태는 $[S' \rightarrow S, \$]$, 최종 상태는 $[S' \rightarrow S, \$]$ 이다. 전이 함수는 \mapsto 로 정의되며, δ_1, δ_2 가 하나 또는 두개의 스택 심볼을 나타내고, α 가 ϵ 이거나 터미널 심볼일 때 $\delta_1 \mapsto \alpha \delta_2$ 형태로 표현된다. 아래와 같은 세 가지 타입의 전이로 구성된다.

- (a) 읽기타입: $[A \rightarrow a.\alpha\beta, u] \mapsto^a [A \rightarrow a\alpha.\beta, u]$
 - (b) 클로져타입: $[A \rightarrow a.B\beta, u] \mapsto [A \rightarrow a.B\beta, u] [B \rightarrow \gamma, v]$
(여기서 $v \in \text{FIRST}_k(\beta u)$ 이다.)
 - (c) 리덕션타입: $[A \rightarrow a.B\beta, u] [B \rightarrow \gamma, v] \mapsto [A \rightarrow aB.\beta, u]$
- 이 오토마톤의 컨피규레이션은 δ 가 스택 스트링(또는 스택의 상위 스트링)이고 x 가 남은 입력 스트링일 때 δx 형태로 주어진다. 입력 스트링 x 에 대한 시작 컨피규레이션은 $[S' \rightarrow S, \$]_1 \#$ 이고 최종 컨피규레이션은 $[S' \rightarrow S, \$]_1 \#$ 이다. $\delta_1 \mapsto \alpha \delta_2$ 의 전이가 있는 경우에

$[S \rightarrow S, \beta]x \Rightarrow \widehat{\pi}[S \rightarrow S, \beta]x$ 인 $\widehat{\pi}$ 가 존재한다. \square

다음은 \widehat{M}_k^c 내의 타당한 동작들의 부분 배열에 대한 $T(G, M)$ 상의 유도 과정을 보이고 있다.

보조 정리 4.1 \widehat{M}_k^c 내에 타당한 동작들의 부분 배열

$[S \rightarrow S, \beta]x \Rightarrow \widehat{\pi}[B \rightarrow \beta, \gamma, \nu]x$ 인 $\widehat{\pi}$ 가 존재한다고

하자. 이때 어떤 $A_j, q^{A_j}, a_j, q_{a_j}^{A_j}, j=1, 2, \dots, m(m \geq 1), A_{m+1} = \epsilon$ 에 대해 $T(G, M)$ 상에 $\tau_T(\pi_T) = \tau_{\widehat{\pi}}(\widehat{\pi})$ 인

$[q^S, \epsilon, T(q^S)] \Rightarrow_{\widehat{\pi}}^{\tau_T} [q^{A_1}, a_1, q_{a_1}^{A_1}] [q^{A_2}, a_2, q_{a_2}^{A_2}] \dots [q^{A_m}, a_m, q_{a_m}^{A_m}] \dots$

$[q^{A_1}, a_1, q_{a_1}^{A_1}]$ 인 π_T 가 존재한다.

(증명) 이 보조 정리에 대한 자세한 증명은 지면상의 제한으로 [3]을 참조하길 바란다. \square

성질 2.2, 성질 4.5, 보조 정리 4.1로부터 다음의 보조 정리를 얻을 수 있다.

보조 정리 4.2 G 상에 $S \Rightarrow_{\widehat{\pi}}^{\tau_T} x$ 인 $\widehat{\pi}$ 가 있다면, $\tau_T(\pi_T) = \widehat{\pi}^R$ 인 $[q^S, \epsilon, T(q^S)] \Rightarrow_{\widehat{\pi}}^{\tau_T} x$ 인 π_T 가 $T(G, M)$ 상에 존재한다. \square

다음은 주어진 $T(G, M)$ 내의 좌측 유도 과정에 대응하는 G 내의 우측 유도 과정을 보인다.

보조 정리 4.3 $T(G, M)$ 상에 $[q^A, a, q_a^A] \Rightarrow_{\widehat{\pi}}^{\tau_T} x$ 인 π_T 가 존재한다면, $\pi = \tau_T(\pi_T)^R$ 인 $A \Rightarrow_{\widehat{\pi}}^{\tau_T} x$ 인 π 가 G 상에 존재한다.

(증명) [2, 보조정리 4.6]에서의 방법과 동일한 방법으로 x 의 길이에 관한 인덕션에 의해 가능하다. 자세한 과정은 언급하지 않는다. \square

보조정리 4.3에서 $A = S, a = \epsilon, q_a^A = T(q^S)$ 으로 선택 시에 다음의 보조 정리를 얻는다.

보조 정리 4.4 $[q^S, \epsilon, T(q^S)] \Rightarrow_{\widehat{\pi}}^{\tau_T} x$ 인 π_T 가 $T(G, M)$ 상에 있다면 $\pi = \tau_T(\pi_T)^R$ 인 $S \Rightarrow_{\widehat{\pi}}^{\tau_T} x$ 인 π 가 G 상에 있다. \square

보조정리 4.2와 보조정리 4.4로부터 $T(G, M)$ 는 G 를 left-to-right 커버한다. 한편 $T(G, M)$ 는 $LL(k)$ 임을 정리 3.1에서 언급했다. 따라서 $T(G, M)$ 는 G 에 대한 $LL(k)$ 커버이다.

정리 4.1 $T(G, M)$ 는 G 에 대한 $LL(k)$ 커버이다. \square

5. 결론

본 논문에선 기존의 LL 커버링 문법들의 생성을 위해 사용된 방법인 LR 파서의 행동 시뮬레이션의 확장을 위한 아이디어와 정형적인 방법을 보였다. 구체적으

로는 기존의 스택 스트링 표현법을 파싱 문맥을 나타내는 LR 아이템을 추가하여 정제하였다. 또한 리덕션 심볼에 대한 예상을 확장하기 위해서 밀방향 전이뿐 아니라 옆방향 전이까지 포함한 정형식을 보였다. 그 결과 행동 시뮬레이션할 수 있는 LR 문법의 범위를 넓힐 수 있었다.

변환 방법이 적용될 수 있는 문법은 확장된 k -transformable 문법으로 정의된다. 어떤 문법이 확장된 k -transformable 문법인지의 여부는 순환적인 너터미널의 존재 유무로 판단되어진다. 따라서 확장된 k -transformable 문법인지의 판단 여부는 결정적이다. 그리고 직관적으로 확장된 k -transformable 문법은 k -transformable 문법과 PLR 문법 클래스를 포함한다. 특히 기존의 LL 커버링 문법이 존재하는 최대의 클래스로 추측되어온 k -transformable 문법의 클래스를 확장 하였기에, 확장된 k -transformable 문법의 의미는 중요하다고 판단된다.

참고 문헌

- [1] Soisalon-Soininen, E. and Ukkonen, E., "A Method for Transforming Grammars into LL(k) Form", pp.338-369, Acta Informatica, Vol 12, 1979.
- [2] Hammer, M., A New Grammatical Transformation into deterministic top down form., p.301, MIT, Mass., Project MAC Technical Report TR-119, 1974.
- [3] Lee, Gyung-Ok and Choe, Kwang-Moo, An Extension of LL(k) Covering Grammars, KAIST, Technical Report, CS-TR-98-133, 1998.
- [4] Aho, A.V. and Ullman, J.D., The Theory of Parsing, Translation and Compiling, vols.1 2., p.1002, Englewood Cliffs, NJ:Prentice-Hall, 1972, 1973.
- [5] Sippu, S. and Soisalon-Soininen, E., Parsing Theory, vol I, II, p.228, p.426, Springer-Veriag Berlin Heidelberg, 1990.

이 경 옥
제 26 권 제 7 호(B) 참조

최 광 무
제 26 권 제 7 호(B) 참조